CrossMark

Convergence Analysis of Difference-of-Convex Algorithm with Subanalytic Data

Hoai An Le Thi[1] (ORCID) **· Van Ngai Huynh[2] · Tao Pham Dinh[3]**

Abstract

Difference-of-Convex programming and related algorithms, which constitute the backbone of nonconvex programming and global optimization, were introduced in 1985 by Pham Dinh Tao and have been extensively developed by Le Thi Hoai An and Pham Dinh Tao since 1994 to become now classic and increasingly popular. That algorithm is a descent method without linesearch and every limit point of its generated sequence is a critical point of the related Difference-of-Convex program. Determining its convergence rate is a challenging problem. Its knowledge is crucial from both theoretical and practical points of view. In this work, we treat this problem for the class of Difference-of-Convex programs with subanalytic data by using the nonsmooth form of the Lojasiewicz inequality. We have successfully proved that the whole sequence is convergent, if it is bounded, provided that the objective function is subanalytic continuous on its domain and one of the two Difference-of-Convex components is differentiable with locally Lipschitz derivative. We also established a result on the convergence rate, which depended on the Lojasiewicz exponent of the objective function. Finally, for both classes of trust-region subproblems and nonconvex quadratic programs, we showed that the Lojasiewicz exponent was one half, and thereby, our proposed algorithms applied to these Difference-of-Convex programs were Root-linearly convergent.

Keywords Difference-of-Convex programming · Difference-of-Convex algorithm · Subdifferential · Subanalyticity · Lojasiewicz exponent · Convergence rate

Mathematics Subject Classification 90C30 · 90C26 · 47N10

Communicated by Jérôme Bolte.

✉ Hoai An Le Thi
hoai-an.le-thi@univ-lorraine.fr
http://lita.univ-lorraine.fr/~lethi

Extended author information available on the last page of the article

⊘ Springer

1 Introduction

Difference-of-Convex functions (DC) programming and DC algorithm (DCA) consti-
tute the backbone of nonconvex programming and global optimization. They play a
key role in these research areas because most real-world nonconvex programs are DC
programs. Based on local optimality conditions and DC duality, DCA is one of rare
effective and efficient algorithms in the nonsmooth nonconvex programming frame-
work. Due to its local character, it cannot guarantee the globality of computed solutions
for general DC programs. However, we observe that, with a suitable starting point,
it converges quite often to a global one. The original key idea of DCA relies on the
DC structure of objective and constraint functions in nonconvex programs, which are
explored and exploited in a deep and suitable way. The resulting DCA introduces the
nice and elegant concept of approximating a nonconvex (DC) program by a sequence
of convex ones: each iteration of DCA requires solution of a convex program. Interest-
ingly, with appropriate DC decompositions, DCA permits to recover, as special cases,
all (resp. most) existing algorithms in convex (resp. nonconvex) programming, e.g.,
the proximal and (sub)gradient algorithms (see [1–6] and references therein).

Being a descent method without linesearch (greatly appreciated in the large-scale
setting), with global convergence (i.e., from an arbitrary starting point), DCA was
successfully applied to lots of nonconvex optimization problems in many fields of
Applied Sciences: Transport Logistics, Telecommunications, Genomics, Finance,
Data Mining-Machine Learning, Cryptology, Computational Biology, Computational
Chemistry, Combinatorial Optimization, Mechanics, Image Processing, Robotics &
Computer Vision, Petrochemicals, Optimal Control and Automatic, Inverse Prob-
lems and Ill-posed Problems, Multiobjective Programming, Game Theory, Variational
Inequalities Problems (VIP), Mathematical Programming with Equilibrium Con-
straints (MPEC), to cite but a few (see the list of references in [5,6]). The popularity of
DC programming and DCA resides in their rich and deep mathematical foundations,
simplicity, flexibility and efficiency of DCAs, their adaptation to specific structures
of addressed problems and their ability to solve real-world large-scale nonconvex
programs.

The purpose of the present paper is to study the convergence of the DCA via the
Łojasiewicz gradient inequality. This celebrated inequality is a key tool in subanalytic
geometry as well as in its applications to subanalytic continuous/discrete dynamical
systems, which has been discovered by Łojasiewicz for differentiable subanalytic
functions, and generalized to nonsmooth subanalytic functions by Bolte–Daniilidis–
Lewis in 2007 [7] . The notion of subanalyticity and its basic properties necessary
for our work will be given in the next section. The class of subanalytic functions,
containing analytic functions, was first introduced by Łojasiewicz [8–10] and then
has been extensively studied by many authors (see, e.g., [7,11,12] and references
therein). It has numerous interesting properties and is rich enough to cover a lot of
practical optimization problems. In [13], the Łojasiewicz subgradient inequality has
been applied to formulate the convergence result for the gradient method. Later, in the
papers [14,15], by using the nonsmooth version of this inequality, the authors have
proved the convergence results of the proximal type methods.

Based on the aforementioned works, we established the convergence result of DCA for DC programming with subanalytic data. It is known (see, e.g., [2–4,16]) that the proximal algorithm in DC programming can be regarded as a special case of DCA; hence, the convergence result in this paper generalizes and completes some results in [13–15].

The rest of the paper is organized as follows. Section 2 presents preliminaries and a background of our motivation: some basic tools of convex and nonsmooth analysis concerning duality, criticality, directional stationarity and local optimality in DC programming, which serves to design the DCA and to prove its sequential convergence. Moreover, the subanalyticity notion, its essential properties, and the Łojasiewicz subgradient inequality are briefly indicated. Section 3 describes the main results of the paper: the convergence rate of the DCA. At last, the final section points out the explicit computation of Łojasiewicz exponents for both classes of trust-region subproblems/quadratic programs. They are all equal to 1/2. It follows the Root-linear convergence of the DCA for these DC programs. To the best of our knowledge, these results seem to be the first in the literature related to the calculation of the Łojasiewicz exponents.

2 Preliminaries

Let us first recall some notions from convex analysis and nonsmooth analysis, which will be needed hereafter [17–19] for the presentation of the nonsmooth Łojasiewicz inequality and the statement of the DCA's convergence rate for DC programs with subanalytic data. In the sequel, the space $X := \mathbb{R}^n$ is equipped with the canonical inner product $\langle \cdot, \cdot \rangle$. Its dual space Y is identified with X itself. By definition, the effective domain of a function $f : X \to \mathbb{R} \cup \{+\infty\}$, denoted by dom f, is

$$\text{dom } f := \{x \in X : f(x) < +\infty\}.$$

It is called proper, iff dom $f \neq \emptyset$. $\mathcal{S}(X)$ represents the set of lower semicontinuous proper functions $f : X \to \mathbb{R} \cup \{+\infty\}$. The open ball with the center $x \in X$ and radius $\varepsilon > 0$ is denoted by $B(x, \varepsilon)$, while the unit ball (i.e., the closed ball with the center at the origin and unit radius) is denoted by B.

Let $\rho \geq 0$ and C be a nonempty and convex set contained in the effective domain of a function $\theta : C \to \mathbb{R}$. One says that the function θ is ρ-convex on C, iff for all x, $u \in C$, $\lambda \in [0, 1]$ one has

$$\theta(\lambda x + (1 - \lambda)u) \leq \lambda(x) + (1 - \lambda)f(u) - \frac{\rho}{2}\lambda(1 - \lambda)\|x - u\|^2.$$

It amounts to saying that $\theta - (\rho/2)\| \cdot \|^2$ is convex on C. The modulus of strong convexity of θ on C, denoted by $\rho(\theta, C)$ or $\rho(\theta)$ if $C = X$, is given by

$$\rho(\theta, C) = \sup\{\rho \geq 0 : \theta - (\rho/2)\| \cdot \|^2 \text{ is convex on } C\}. \tag{1}$$

Clearly, the ρ-convexity implies the convexity. One says that θ is *strongly convex* on C, iff $\rho(\theta, C) > 0$. The supremum of all $\rho \geq 0$ such that the above inequality is verified is called the convexity modulus of f, while it is strongly convex if $\rho > 0$.

For a proper function $f : X \to \mathbb{R} \cup \{+\infty\}$, the Fréchet subdifferential of f at $x \in \mathrm{dom}\, f$ is defined by

$$\partial^F f(x) = \left\{ y \in Y : \liminf_{h \to 0} \frac{f(x+h) - f(x) - \langle y, h \rangle}{\|h\|} \geq 0 \right\}.$$

For $x \notin \mathrm{dom}\, f$, we set $\partial^F f(x) = \emptyset$. If f is a convex function, then $\partial^F f$ coincides with the subdifferential ∂f in the sense of convex analysis. A point $x_0 \in X$ is called a (Fréchet) *critical* (or stationary) point for the function f, if $0 \in \partial^F f(x_0)$.

If f is locally Lipschitz at $x \in X$, then the Clarke directional derivative $f^C(x; .)$ at x and the Clarke subdifferential $\partial^C f(x)$ of f at x are defined by

$$f^C(x; d) := \lim_{(t,u) \to (0^+, x)} \frac{f(u + td) - f(u)}{t},$$

and

$$\partial^C f(x) := \{ y \in Y : \langle y, d \rangle \leq f^C(x; d), \forall d \in X \}.$$

If f is continuously differentiable at x, then $\partial^C f(x) = \{ \nabla f(x) \}$ (the Fréchet derivative of f at x). When f is a convex function, then $\partial^C f$ coincides with the subdifferential ∂f.

For given two locally Lipschitz functions f, g at a given $x \in X$, one has

$$\partial^C (-f)(x) = -\partial^C f(x), \quad \partial^C (f + g)(x) \subset \partial^C f(x) + \partial^C g(x),$$

where the equality in the latter inclusion holds if f is continuously differentiable at x.

Moreover, given a DC-function f, i.e., $f := g - h$, where g and h are convex functions, if h is continuous at x, then

$$\partial^F f(x) \subset \partial g(x) - \partial h(x).$$

Especially, if h is differentiable at x, then one has the equality:

$$\partial^F f(x) = \partial g(x) - \nabla h(x).$$

We now briefly recall the notion of subanalytic sets and functions and some basic properties [8–10].

Definition 2.1 (i) A subset C of X is said to be semianalytic, iff each point of X admits a neighborhood V such that

$$C \cap V = \bigcup_{i=1}^{p} \bigcap_{j=1}^{q} \{ x \in V : f_{ij}(x) = 0, \ g_{ij}(x) > 0 \},$$

where f_{ij}, $g_{ij} : V \to \mathbb{R}$ $(1 \leq i \leq p, 1 \leq j \leq q)$ are real-analytic functions.

(ii) A subset C of X is called subanalytic, iff each point of X admits a neighborhood V such that

$$C \cap V = \{x \in X : \exists y \in \mathbb{R}^m, (x, y) \in D\},$$

where D is a bounded semianalytic subset of $X \times \mathbb{R}^m$ with $m \geq 1$.

(iii) A function $f : X \to \mathbb{R} \cup \{+\infty\}$ is subanalytic, iff its graph is a subanalytic subset of $X \times \mathbb{R}$.

Evidently, the class of subanalytic sets (resp. functions) contains all analytic sets (resp. functions). Subanalytic sets and subanalytic functions enjoy interesting properties. For instance, the class of subanalytic sets is closed under locally finite unions/intersections, relative complements, and the usual projection. The distance function to a subanalytic set is subanalytic; the sum/difference of continuous and subanalytic functions is also subanalytic. For further properties of subanalytic set/functions, the readers are referred to [9,11,12] and the references therein.

The following proposition gives the subanalyticity of the conjugate of strongly convex functions.

Proposition 2.1 *If $f : X \to \mathbb{R} \cup \{+\infty\}$ is a lower semicontinuous, subanalytic and strongly convex function, then its conjugate f^* is a $C^{1,1}$ (the class of functions whose derivatives are Lipschitz), subanalytic, and convex function.*

Proof It is well known that the conjugate of a lower semicontinuous and strongly convex function is of the class $C^{1,1}$. It suffices to show that f^* is subanalytic. Let $\rho > 0$ be the modulus of the strong convexity of f. Take $\rho_1 \in \,]0, \rho[$ and define $h(x):=f(x) - \frac{\rho_1}{2}\|x\|^2$, $x \in X$. Then, h is strongly convex. By the definition of the conjugate,

$$
\begin{aligned}
f^*(y) &= -\inf_{x \in X}\{h(x) + \tfrac{\rho_1}{2}\|x\|^2 - \langle x, y \rangle\} \\
&= -\inf_{y \in \mathbb{R}^n}\{h(y) + \tfrac{\rho_1}{2}\|x/\rho_1 - y\|^2\} + \tfrac{1}{2\rho_1}\|x\|^2 \\
&= -\varphi(y) + \tfrac{1}{2\rho_1}\|y\|^2,
\end{aligned}
$$

with

$$\varphi(y) = \inf_{x \in X}\left\{h(x) + \frac{\rho_1}{2}\left\|x - \frac{y}{\rho_1}\right\|^2\right\}, \quad y \in Y.$$

According to Proposition 2.9 in [7], the function φ is subanalytic. Hence, f^*, being the difference of two subanalytic functions, is subanalytic too. $\qquad\square$

Let us recall next the Łojasiewicz subgradient inequality, established by Bolte–Daniilidis–Lewis [7], which plays a key role in our convergence analysis of the DCA.

Theorem 2.1 (Theorem 3.1, [7]) *Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a subanalytic function such that: its domain $\mathrm{dom}\, f$ is closed; $f \mid_{\mathrm{dom}\, f}$ is continuous. Let x_0 be a Fréchet*

critical point of f. Then, there exist $\theta \in [0, 1[$, $L > 0$ and a neighborhood V of x_0 such that

$$|f(x) - f(x_0)|^\theta \le L\|x^*\| \text{ for all } x \in V, \ x^* \in \partial^F f(x),$$

where the convention $0^0 = 1$ is used.

The number θ in the theorem is called a *Łojasiewicz exponent* of the function f at the critical point x_0.

3 Convergence Analysis of DCA

This main section is devoted to the statement of DCA's convergence rate for DC programs with subanalytic data. It requires some basic tools of DC programming and DCA indispensable for achieving this objective.

Introduction to DC Programming and DCA

Let $\Gamma_0(X)$ denote the convex cone of all lower semicontinuous, proper, and convex functions on X. The indicator function χ_C of a nonempty, closed, and convex set $C \subset X$ is defined by $\chi_C(x) = 0$, if $x \in C$, $+\infty$, otherwise. One has $\chi_C \in \Gamma_0(X)$. The vector space of DC functions is denoted by $DC(X) := \Gamma_0(X) - \Gamma_0(X)$. It is well known that this class of DC functions is quite large to contain almost real-life objective functions, and is closed under all the usual operations encountered in optimization.

We consider the standard DC program of the form:

$$\alpha := \inf\{f(x) := g(x) - h(x) : x \in X\} \qquad (P_{dc})$$

with $g, h \in \Gamma_0(X)$. The function f is called a DC-function on X, while g and h are its DC components. By replacing the convex functions g and h with $g + \frac{\rho}{2}\|\cdot\|^2$ and $h + \frac{\rho}{2}\|\cdot\|^2$, ρ being a positive scalar, one can, without loss of generality, assume the strong convexity of both functions g and h in (P_{dc}).

Clearly, a DC-function f has infinitely many DC decompositions (each of which gives rises to its own DCA), which could strongly impact on practical performances (optimality, quality of computed solutions, convergence rate, efficiency, and scalability, etc) of the DCA [2–6,20,21]. Note that any closed and convex constraint set $C \subset \text{dom } g$ can be incorporated in the objective function of (P_{dc}) by using the indicator function χ_C:

$$\inf\{f(x) := g(x) - h(x) : x \in C\} = \inf\{\chi_C(x) + g(x) - h(x) : x \in X\}. \qquad (2)$$

From the literature of subanalytic geometry, under some suitable conditions, the function $f + \chi_C$ is subanalytic, if both f and C are subanalytic. Recall that, under the usual convention $+\infty - (+\infty) = +\infty$, the finiteness of the optimal value α implies

$$\text{dom } f = \text{dom } (g + \chi_C) = \text{dom } g \cap \text{dom } \chi_C = C \subset \text{dom } h.$$

The dual DC program (D_{dc}) of (P_{dc}) is defined by

$$\alpha = \inf\{h^*(y) - g^*(y) : y \in Y\}. \qquad (D_{dc})$$

Here g^* and h^* are the conjugate functions of g and h, respectively, i.e.,

$$g^*(y):= \sup\{\langle x, y \rangle - g(x) : x \in X\}.$$

Critical and strongly critical points of $g - h$

A point x^* is a critical point of (P_{dc}) (or $f = g - h$), iff $\partial g(x^*) \cap \partial h(x^*) \neq \emptyset$, or equivalently $0 \in \partial g(x^*) - \partial h(x^*)$, while it is called strongly critical point of (P_{dc}) (or $f = g - h$), iff $\emptyset \neq \partial h(x^*) \subset \partial g(x^*)$. The sets of critical (resp. strongly critical) points of $g - h$ are denoted by $\mathcal{P}_c$ (resp. $\mathcal{P}_{sc}$), and $\mathcal{P}_l$ is the set of local minimizers of $g - h$. Likewise, replacing (P_{dc}) by (D_{dc}) gives the similar notations for $h^* - g^*$.

Necessary local optimality conditions for the primal DC program (P_{dc}) were developed as follows (by symmetry those relating to dual DC program (D_{dc}) are trivially deduced) [2–5].

The notion of DC criticality is close to Clarke/Fréchet stationarity in the following sense:

First, by definition, x^* is Clarke/Fréchet subdifferential $\partial^C f$ (resp. $\partial^F f$) of $f = g - h$, iff

$$0 \in \partial^C f(x^*) \qquad (3)$$

(resp.

$$0 \in \partial^F f(x^*)). \qquad (4)$$

On the other hand, it is well known that

$$\partial^C f(x) \subset [\partial g(x) - \partial h(x)],$$
$$\partial^F f(x) \subset [\partial g(x) - \partial h(x)] \qquad (5)$$

with equality under technical assumptions. Hence, Clarke stationarity of x^* or its Fréchet stationarity implies *DC criticality* of x^*. Their equivalence occurs if the related equality holds in the corresponding inclusion.

DC Strong Criticality and Directional Stationarity for DC Programs

As remarked above, DC criticality and strong criticality depend on DC decompositions $g - h$ of DC objective function $f = g - h$. Since a DC-function has infinitely many DC decompositions, it is important to find relations between DC criticality and the commonly used directional stationarity (*d-stationarity*, for short) defined below, which depends only on f but not on DC components g and h.

For the sake of completeness, let us recall and complete the major results concerning these different criticalities. They rely on some main results, in modern convex analysis [18,22], related to the subdifferential, the directional derivative of a convex function,

and the support function of a convex set in X. Let $\varphi : X \to \mathbb{R} \cup \{+\infty\}$ be a proper function on X and $x \in \mathrm{dom}\, \varphi$.

The directional derivative $\varphi'(x; \cdot)$ of φ at x is defined by

$$\varphi'(x; u) := \lim_{t \downarrow 0} \frac{\varphi(x + tu) - \varphi(x)}{t}.$$

If φ is convex, then it becomes

$$\varphi'(x; u) := \inf_{t > 0} \frac{\varphi(x + tu) - \varphi(x)}{t}.$$

Let us recall the following relations between the subdifferential $\partial \varphi(x)$ of φ at x and the directional derivative $\varphi'(x; \cdot)$ of φ at x:

Assume that the function $\varphi : X \to \mathbb{R} \cup \{+\infty\}$ is proper convex on X, then for $x \in \mathrm{dom}\, \varphi$:

a) $\varphi'(x; u)$ is convex and positively homogeneous as a function of u,
b) $y \in \partial \varphi(x)$ iff $\varphi'(x; u) \geq \langle u, y \rangle$ for every $u \in X$, i.e.,

$$\varphi'(x; u) \geq \sup\{\langle u, y \rangle : y \in \partial \varphi(x)\} = \chi^*_{\partial \varphi(x)}(u).$$

In fact, the closure of $\varphi'(x; u)$, as a convex function of u, is the support function of the closed and convex set $\partial \varphi(x)$

$$\mathrm{cl}(\varphi'(x; u)) = \chi^*_{\partial \varphi(x)}(u).$$

Moreover, if $x \in \mathrm{ri}(\mathrm{dom}\, \varphi)$, then $\partial \varphi(x)$ is nonempty and $\varphi'(x; u)$ is semicontinuous and proper on X as a function of u such that

$$\varphi'(x; u) = \sup\{\langle u, y \rangle : y \in \partial \varphi(x)\} = \chi^*_{\partial \varphi(x)}(u).$$

Let C be a nonempty, closed, and convex set in X. Then, for $x \in C$,

$$\partial \chi_C(x) = \{y \in Y : \langle u - x, y \rangle \leq 0 \ \forall u \in C\} = N(C, x),$$

where $N(C, x)$ is the (closed and convex) normal cone to C at x.

c) Let C_1 and C_2 be two closed and convex sets in X. Then,

$$C_1 \subset C_2 \Leftrightarrow \chi^*_{C_1} \leq \chi^*_{C_2}.$$

The d-Stationarity of a Proper Function $\varphi : X \to \mathbb{R} \cup \{+\infty\}$

A vector $x \in \mathrm{dom}\, \varphi$ is *d-stationary* of φ, iff $\varphi'(x; u) \geq 0$ for every $u \in X$.

The following result shows the key relations between *d-stationarity* and *strong criticality*, whose proof is a direct consequence of the properties a), b), and c) aforementioned, and is omitted here.

Theorem 3.1 *Let $f := g - h$ be the DC objective function of the primal DC program* (P_{dc}) *whose optimal value α is finite. Then, the following properties hold:*

(i) *$f'(x; u) = g'(x; u) - h'(x; u)$ for all $u \in X$.*

(ii) *The vector $x \in \mathrm{dom}\, f$ is d-stationary of f (or for (P_{dc})), iff $g'(x; u) \geq h'(x; u)$ for all $u \in X$. Hence, for $x \in \mathrm{dom}\, f$, if both equalities $g'(x; u) = \chi^*_{\partial g(x)}(u)$ and $h'(x; u) = \chi^*_{\partial h(x)}(u)$ hold for all $u \in X$, then there is an identity between the d-stationarity of x and the strong criticality of x.*

(iii) *If $x \in \mathrm{ri}(\mathrm{dom}\, g) \cap \mathrm{ri}(\mathrm{dom}\, h)$, then the d-stationarity of x is equivalent to the strong criticality of x.*

(iv) *If $x \in \mathrm{ri}(\mathrm{dom}\, g)$, then the d-stationarity of x implies the strong criticality of x.*

(v) *If $x \in \mathrm{ri}(\mathrm{dom}\, h)$, then the strong criticality of x implies the d-stationarity of x.*

(vi) *If the affine hulls of $\mathrm{dom}\, g$ and of $\mathrm{dom}\, h$ are identical, then $\mathrm{ri}(\mathrm{dom}\, g) \subset \mathrm{ri}(\mathrm{dom}\, h)$, and for $x \in \mathrm{ri}(\mathrm{dom}\, g)$, the d-stationarity of x is equivalent to the strong criticality of x.*

Duality and Local Optimality Conditions in DC Programming

Let us summarize some *basic properties about duality and local optimality* in DC programming [2–6] that will be used in the following for the construction of the DCA.

Theorem 3.2 (i) *If x^* is a local minimizer of $g - h$, then $x^* \in \mathcal{P}_\ell$.*

(ii) *Let x^* be a critical point of $g - h$ and $y^* \in \partial g(x^*) \cap \partial h(x^*)$. Let U be a neighborhood of x^* such that $(U \cap \mathrm{dom}\, g) \subset \mathrm{dom}\, \partial h$. If for any $x \in U \cap \mathrm{dom}\, g$ there is $y \in \partial h(x)$ such that $h^*(y) - g^*(y) \geq h^*(y^*) - g^*(y^*)$, then x^* is a local minimizer of $g - h$. More precisely,*

$$g(x) - h(x) \geq g(x^*) - h(x^*), \forall x \in U \cap \mathrm{dom}\, g.$$

Corollary 3.1 (Sufficient local optimality) *Let x^* be a point that admits a neighborhood U such that $\partial h(x) \cap \partial g(x^*) \neq \emptyset, \forall x \in U \cap \mathrm{dom}\, g$. Then, x^* is a local minimizer of $g - h$. More precisely,*

$$g(x) - h(x) \geq g(x^*) - h(x^*), \forall x \in U \cap \mathrm{dom}\, g.$$

Corollary 3.2 (Sufficient strict local optimality) *If $x^* \in \mathrm{int}(\mathrm{dom}\, h)$ verifies $\partial h(x^*) \subset \mathrm{int}(\partial g(x^*))$, then x^* is a strict local minimizer of $g - h$.*

Corollary 3.3 (DC duality transportation of a local minimizer) *Let $x^* \in \mathrm{dom}\, \partial h$ be a local minimizer of $g - h$ and let $y^* \in \partial h(x^*)$ (i.e., $\partial h(x^*)$ is nonempty, and x^* admits a neighborhood U such that*

$$g(x) - h(x) \geq g(x^*) - h(x^*), \forall x \in U \cap \mathrm{dom}\, g). \tag{6}$$

If

$$y^* \in \mathrm{int}(\mathrm{dom}\, g^*) \text{ and } \partial g^*(y^*) \subset U \tag{7}$$

((7) *holds if g^* is differentiable at y^*), then y^* is a local minimizer of $h^* - g^*$.*

DCA Scheme and Its Convergence Properties

Philosophy of DCA The DCA is based on local optimality conditions and duality in DC programming. The main idea of DCA is quite simple: at each iteration k, DCA approximates the second DC component h by its affine minorization $h_k(x):=h(x^k)+ \langle x - x^k, y^k \rangle$, with $y^k \in \partial h(x^k)$, and minimizes the resulting convex function.

Basic DCA scheme
Initialization: Let $x^0 \in \text{dom } \partial h$.
For $k = 0, 1, \ldots$ until convergence of $\{x^k\}$:

Step 1: Compute $y^k \in \partial h(x^k)$.
Step 2: Compute $x^{k+1} \in \text{argmin}\{g(x) - h_k(x) : x \in X\}$ (P_k).

Recall the core results on DCA's convergence [2–5].

Theorem 3.3 *The DCA is a descent method without linesearch, but with global convergence, which enjoys the following properties: (C and D are two convex sets in X, containing the sequences $\{x^k\}$ and $\{y^k\}$ respectively)*

(i) *The sequences $\{g(x^k) - h(x^k)\}$ and $\{h^*(y^k) - g^*(y^k)\}$ are decreasing and convergent to the same limit, and*

- $g(x^{k+1}) - h(x^{k+1}) = g(x^k) - h(x^k)$ *iff* $[\rho(g, C) + \rho(h, C)]\|x^{k+1} - x^k\| = 0$, $y^k \in \partial g(x^k) \cap \partial h(x^k)$, *and* $y^k \in \partial g(x^{k+1}) \cap \partial h(x^{k+1})$. *Moreover if g or h is strictly convex on C, then $x^k = x^{k+1}$. In such a case, the DCA terminates at the kth iteration (finite convergence of DCA).*
- $h^*(y^{k+1}) - g^*(y^{k+1}) = h^*(y^k) - g^*(y^k)$ *iff* $x^{k+1} \in \partial g^*(y^k) \cap \partial h^*(y^k)$, $x^{k+1} \in \partial g^*(y^{k+1}) \cap \partial h^*(y^{k+1})$, *and* $[\rho(g^*, D) + \rho(h^*, D)]\|y^{k+1} - y^k\| = 0$. *Moreover if g^* or h^* is strictly convex on D, then $y^{k+1} = y^k$. In such a case, the DCA terminates at the kth iteration (finite convergence of DCA).*

(ii) *If $\rho(g, C) + \rho(h, C) > 0$ (resp. $\rho(g^*, D) + \rho(h^*, D) > 0$), then the series $\{\|x^{k+1} - x^k\|^2\}$ (resp. $\{\|y^{k+1} - y^k\|^2\}$) is convergent.*
(iii) *If the optimal value α of problem (P_{dc}) is finite and the sequences $\{x^k\}$ and $\{y^k\}$ are bounded, then every limit point $\tilde{x}$ (resp. $\tilde{y}$) of the sequence $\{x^k\}$ (resp. $\{y^k\}$) is a critical point of $g - h$ (resp. $h^* - g^*$).*
(iv) *For polyhedral DC program (i.e., a DC program in which at least one of the functions g and h is polyhedral), the sequences $\{x^k\}$ and $\{y^k\}$ contain finitely many iterates and hence they have finite convergence (i.e., after a finite number of iterations).*

Remark 3.1 By using the optimal choice of subgradients [2–4] in the construction of $\{x^k\}$ and $\{y^k\}$, the modified DCA (called the complete DCA in the works just mentioned above) sequentially converges to strongly critical points of (P_{dc}) and (D_{dc}), respectively. The convergence result for the DCA is stated in the following theorem. The main idea of its proof, inspired from [13,14], is based on the Łojasiewicz original inequality (see also [15] in which the Kurdyka-Łojasiewicz inequality, a general version of the Łojasiewicz one, is used). It can be described roughly as follows. Given

a sequence $\{x^k\}$ defined by the DCA, the series $\sum_{k=0}^{\infty} \|x^k - x^{k+1}\|^2$ is convergent according to Theorem 3.3. The help of the Łojasiewicz subgradient inequality is to guarantee the convergence of the series $\sum_{k=0}^{\infty} \|x^k - x^{k+1}\|$ and, as a result, the convergence rate of DCA for DC programs with subanalytic data.

Theorem 3.4 *Let $\{x^k\}$ and $\{y^k\}$ be two sequences generated by the DCA. Then the following properties hold:*

(i) *Suppose that the DC-function $f := g - h$ is subanalytic; dom f is closed; $f \mid_{\mathrm{dom} f}$ is continuous; and around every critical point of (P_{dc}), either g or h is differentiable with locally Lipschitz derivative. Assume that $\rho := \rho(g) + \rho(h) > 0$. If either the sequence $\{x^k\}$ or $\{y^k\}$ is bounded, then $\{x^k\}$ and $\{y^k\}$ are convergent to critical points of (P_{dc}) and (D_{dc}), respectively.*
(ii) *Similarly, the dual problem has its counterpart: suppose that $h^* - g^*$ is subanalytic; $\mathrm{dom}(h^* - g^*)$ is closed; $(h^* - g^*) \mid_{\mathrm{dom}(h^*-g^*)}$ is continuous; and around critical points of (D_{dc}), either g^* or h^* is differentiable with locally Lipschitz derivative. If $\rho(g^*) + \rho(h^*) > 0$ and either the sequence $\{x^k\}$ or $\{y^k\}$ is bounded, then $\{x^k\}$ and $\{y^k\}$ are convergent to critical points of (P_{dc}) and (D_{dc}), respectively.*

Proof In view of the DC duality in Theorem 3.3 [2–6], it suffices to prove the statement (i). Due to Proposition 2.1, either g^* or h^* is of the class $C^{1,1}$. Since $x^{k+1} \in \partial g^*(y^k)$ and $x^k \in \partial h^*(y^k)$, then $\{x^k\}$ is bounded iff $\{y^k\}$ is bounded. Moreover, since either g or h is differentiable, $\{x^k\}$ is convergent iff $\{y^k\}$ is convergent. Thus, without loss of generality, we can assume that $\{x^k\}$ is bounded.

If $x^{k+1} = x^k$ for some k, then x^k is a critical point of (P_{dc}) as shown in Theorem 3.3. Assume now that $x^{k+1} \neq x^k$ for all k. Also from Theorem 3.3, one has

$$f(x^k) - f(x^{k+1}) \geq \frac{\rho}{2} \|x^k - x^{k+1}\|^2, \quad \text{for } k = 0, 1, \ldots, \tag{8}$$

and any limit point of $\{x^k\}$ is a critical point of $(g - h)$.

Inequality (8) implies immediately that $\{f(x^k)\}$ is decreasing, and as $\{x^k\}$ is a bounded sequence, yielding the convergence of the sequence $\{f(x^k)\}$. Therefore, $\lim_{k \to \infty} \|x^k - x^{k+1}\| = 0$. Considering the function $f - \lim_{k \to \infty} f(x^k)$ instead of f, we can assume that $\lim_{k \to \infty} f(x^k) = 0$. Then, $f(x^k) > 0$ for all $k \in \mathbb{N}$. Denote by $\mathcal{C}$ the set of all limit points of $\{x^k\}$. Then, every point belonging to $\mathcal{C}$ is a critical point of (P_{dc}), and moreover, the boundedness of the sequence $\{x^k\}$ implies that the set $\mathcal{C}$ is compact.

For each $x \in \mathcal{C}$, by our assumption, there exist $\kappa(x), \varepsilon(x) > 0$ such that one of the following two cases occurs:

– g is differentiable on $B(x, \varepsilon(x))$ and

$$\|\nabla g(u) - \nabla g(v)\| \leq \kappa(x)\|u - v\| \quad \forall u, v \in B(x, \varepsilon(x)). \tag{9}$$

In this case, x is a Fréchet critical point of the function $(-f)$, and since $\mathrm{dom} g \subset \mathrm{dom} h$, $(-f)$ is finite, continuous, and subanalytic on $B(x, \varepsilon(x))$. According to

the Łojasiewicz inequality applied to the subanalytic function $-f$, by shrinking $\varepsilon(x)$ if necessary, we can find $L(x) > 0$ and $\theta(x) \in [0, 1[$ such that

$$|f(u)|^{\theta(x)} \leq L(x)\|u^*\| \quad \text{for all } u \in B(x, \varepsilon(x)), u^* \in \partial^F(-f)(u). \quad (10)$$

– h is differentiable on $B(x, \varepsilon(x))$ and

$$\|\nabla h(u) - \nabla h(v)\| \leq \kappa(x)\|u - v\| \quad \forall u, v \in B(x, \varepsilon(x)). \quad (11)$$

Applying the Łojasiewicz inequality to the subanalytic function f around the Fréchet critical point x yields the existence of $L(x) > 0$ and $\theta(x) \in [0, 1[$ such that

$$|f(u)|^{\theta(x)} \leq L(x)\|u^*\| \quad \text{for all } u \in B(x, \varepsilon(x)), \ u^* \in \partial^F f(u). \quad (12)$$

Due to the compactness of $\mathcal{C}$, there are $z^1, z^2, \ldots, z^p \in \mathcal{C}$ such that

$$\mathcal{C} \subset \bigcup_{i=1}^{p} B(z^i, \varepsilon(z^i)/2).$$

Relabeling if necessary, assume, without loss of generality, that

$$x^k \in \bigcup_{i=1}^{p} B(z^i, \varepsilon(z^i)/2) \quad \text{and} \quad \|x^k - x^{k+1}\| < \varepsilon/2, \ k = 0, 1, \ldots,$$

where $\varepsilon = \min\{\varepsilon(z^i)/2 : i = 1, \ldots, p\}$.

Define

$$\theta = \max\{\theta(z^i) : i = 1, \ldots, p\}, \ \kappa = \min\{\kappa(z^i) : i = 1, \ldots, p\},$$
$$L = \min\{L(z^i) : i = 1, \ldots, p\}.$$

For each $k = 0, 1, \ldots$, let $i_k \in \{1, \ldots, p\}$ such that $x^k \in B(z^{i_k}, \varepsilon(z^{i_k})/2)$. By this, $x^{k+1} \in B(z^{i_k}, \varepsilon(z^{i_k}))$. Let us consider the following cases:

Case 1 g is differentiable and its derivative is locally Lipschitz on $B(z^{i_k}, \varepsilon(z^{i_k}))$. From (9) and (10), we derive that

$$\begin{aligned} \|\nabla g(x^k) - \nabla g(x^{k+1})\| &\leq \kappa\|x^k - x^{k+1}\|, \\ |f(x^k)|^{\theta} &\leq L\|x^{k*}\| \quad \text{for } k = 0, 1, \ldots, \ x^{k*} \in \partial^F(-f)(x^k). \end{aligned} \quad (13)$$

As g is differentiable on $B(z^{i_k}, \varepsilon(z^{i_k}))$ and by the definition of $\{x^k\}$, there holds

$$\nabla g(x^{k+1}) - \nabla g(x^k) = y^k - \nabla g(x^k) \in \partial h(x^k) - \nabla g(x^k) = \partial^F(-f)(x^k).$$

Hence, according to (13), one gets

$$|f(x^k)|^\theta \le L\|\nabla g(x^k) - \nabla g(x^{k+1})\| \le L\kappa\|x^k - x^{k+1}\| \quad \text{for all } k = 0, 1, \ldots \tag{14}$$

Thanks to the concavity of the function $t \in \mathbb{R} \mapsto t^{1-\theta}$ on $]0, +\infty[$, and by $f(x^k) > 0$ together with relations (8) and (14), one has the estimate

$$
\begin{aligned}
f(x^k)^{1-\theta} &- f(x^{k+1})^{1-\theta} \\
&\ge (1-\theta)f(x^k)^{-\theta}(f(x^k) - f(x^{k+1})) \\
&\ge \frac{(1-\theta)}{L\kappa\|x^k - x^{k+1}\|}\frac{\rho}{2}\|x^k - x^{k+1}\|^2 = \frac{(1-\theta)\rho}{2L\kappa}\|x^k - x^{k+1}\|.
\end{aligned}
\tag{15}
$$

Case 2 The function h is differentiable and its derivative is locally Lipschitz on $B(z^{i_k}, \varepsilon(z^{i_k}))$.

Then, according to (11) and (12), there hold

$$
\begin{aligned}
\|\nabla h(x^k) - \nabla h(x^{k-1})\| &\le \kappa\|x^k - x^{k-1}\|, \\
|f(x^k)|^\theta &\le L\|x^{k*}\| \text{ for all } k = 0, 1, \ldots, x^{k*} \in \partial^F f(x^k).
\end{aligned}
\tag{16}
$$

Since

$$\nabla h(x^{k-1}) - \nabla h(x^k) = y^{k-1} - \nabla h(x^k) \in \partial g(x^k) - \nabla h(x^k) = \partial^F f(x^k),$$

the relations (16) give

$$|f(x^k)|^\theta \le L\|\nabla h(x^k) - \nabla h(x^{k-1})\| \le L\kappa\|x^k - x^{k-1}\| \quad \text{for all } k = 0, 1, \ldots \tag{17}$$

Similarly, it follows from (15) that

$$
\begin{aligned}
f(x^k)^{1-\theta} - f(x^{k+1})^{1-\theta} &\ge (1-\theta)f(x^k)^{-\theta}(f(x^k) - f(x^{k+1})) \\
&\ge \frac{(1-\theta)\rho\|x^k - x^{k+1}\|^2}{2L\kappa\|x^k - x^{k-1}\|}.
\end{aligned}
$$

Therefore, the inequality $a \le \frac{a^2}{b} + b/4$ for any $a, b > 0$ implies that

$$
\begin{aligned}
\|x^k - x^{k+1}\| &\le \frac{\|x^k - x^{k+1}\|^2}{\|x^k - x^{k-1}\|} + \|x^k - x^{k-1}\|/4 \\
&\le \frac{2L\kappa}{\rho(1-\theta)}\left(f(x^k)^{1-\theta} - f(x^{k+1})^{1-\theta}\right) + \|x^k - x^{k-1}\|/4. \tag{18}
\end{aligned}
$$

In both cases, the formulas (15) and (18) show that for all $k = 1, 2, \ldots,$

$$\|x^k - x^{k+1}\| \leq \|x^k - x^{k-1}\|/4 + \frac{2L\kappa}{\rho(1-\theta)}\left(f(x^k)^{1-\theta} - f(x^{k+1})^{1-\theta}\right).$$

Consequently, by adding these inequalities, one has

$$\sum_{k=1}^{\infty} \|x^k - x^{k+1}\| \leq \frac{\|x^1 - x^0\|}{3} + \frac{8L\kappa}{3\rho(1-\theta)}f(x^0)^{1-\theta},$$

which yields the convergence of the sequence $\{x^k\}$, and the proof is completed. $\quad\square$

Corollary 3.4 *Suppose that $g - h$ and $h^* - g^*$ are subanalytic functions with closed domains such that $(g-h)\,|_{\mathrm{dom}(g-h)}$ and $(h^*-g^*)\,|_{\mathrm{dom}(h^*-g^*)}$ are continuous. Assume that $\rho(g) + \rho(h) > 0$ and $\rho(g^*) + \rho(h^*) > 0$. If either the sequence $\{x^k\}$ or $\{y^k\}$ is bounded, then these sequences converge to critical points of (P_{dc}) and (D_{dc}), respectively.*

Proof Since the conjugate of a strongly convex function is differentiable with Lipschitz derivative, the conclusion follows directly from Theorem 3.4. $\quad\square$

The next theorem provides the convergence rate of the sequence $\{x^k\}$ that depends on the Łojasiewicz exponent of the objective function at its limit point.

Theorem 3.5 *Suppose that the assumptions of* Theorem *3.4 (i) are satisfied. Let x^∞ be the limit point of $\{x^k\}$ at which the Łojasiewicz exponent $\theta \in [0, 1[$ of the function f is given. Then, there exist constants $\tau_1, \tau_2 > 0$ such that*

$$\|x^k - x^\infty\| \leq \sum_{j=k}^{\infty} \|x^j - x^{j+1}\| \leq \tau_1 \|x^k - x^{k-1}\| + \tau_2 \|x^k - x^{k-1}\|^{\frac{1-\theta}{\theta}}, k = 1, 2, \ldots$$

$$\tag{19}$$

As a result, one has

- *If $\theta \in]1/2, 1[$, then $\|x^k - x^\infty\| \leq ck^{\frac{1-\theta}{1-2\theta}}$ for some $c > 0$.*
- *If $\theta \in]0, 1/2]$, then $\|x^k - x^\infty\| \leq cq^k$ for some $c > 0$ and $q \in]0, 1[$.*
- *If $\theta = 0$, then $\{x^k\}$ is convergent in a finite number of iterations.*

Proof The first inequality in (19) is obvious. For the second, set

$$r_k = \sum_{j=k}^{\infty} \|x^j - x^{j+1}\|, \; k = 0, 1, \ldots$$

If g is differentiable with Lipschitz derivative, then by virtue of (14) and (15), one has

$$r_k = \sum_{j=k}^{\infty} \|x^j - x^{j+1}\| \le \frac{2L\kappa}{\rho(1-\theta)} f(x^k)^{1-\theta} \le \frac{2L\kappa}{\rho(1-\theta)} f(x^{k-1})^{1-\theta}$$

$$\le \frac{2(L\kappa)^{\frac{1}{\theta}}}{(1-\theta)\rho} \|x^k - x^{k-1}\|^{\frac{1-\theta}{\theta}}.$$

Suppose now that h is differentiable with Lipschitz derivative. Relations (17) and (18) imply

$$r_k \le \frac{r_k + \|x^k - x^{k-1}\|}{4} + \frac{2L\kappa}{(1-\theta)\rho} f(x^k)^{1-\theta}$$

$$\le \frac{r_k + \|x^k - x^{k-1}\|}{4} + \frac{2(L\kappa)^{\frac{1}{\theta}}}{(1-\theta)\rho} \|x^k - x^{k-1}\|^{\frac{1-\theta}{\theta}}.$$

Hence,

$$r_k \le \frac{\|x^k - x^{k-1}\|}{3} + \frac{8(L\kappa)^{\frac{1}{\theta}}}{3(1-\theta)\rho} \|x^k - x^{k-1}\|^{\frac{1-\theta}{\theta}}$$

and (19) is proved. □

As $\lim_{k\to\infty} \|x^k - x^{k+1}\| = 0$, we can assume that $\|x^k - x^{k+1}\| < 1$ for all $k = 0, 1, \ldots$ By setting $\tau = \tau_1 + \tau_2$ in (19), one has

- If $\theta \in]1/2, 1[$, then $r_k \le \tau(r_{k-1} - r_k)^{\frac{1-\theta}{\theta}}$. Hence, by the convexity of the function $t \mapsto t^{\frac{1-2\theta}{1-\theta}}$ on $]0, +\infty[$,

$$r_k^{\frac{1-2\theta}{1-\theta}} - r_{k-1}^{\frac{1-2\theta}{1-\theta}} \ge \frac{1-2\theta}{1-\theta} r_k^{\frac{-\theta}{1-\theta}} (r_k - r_{k-1}) \ge \frac{2\theta - 1}{(1-\theta)\tau}.$$

Thus,

$$r_k^{\frac{1-2\theta}{1-\theta}} = r_0^{\frac{1-2\theta}{1-\theta}} + \sum_{j=1}^{k} \left(r_k^{\frac{1-2\theta}{1-\theta}} - r_{k-1}^{\frac{1-2\theta}{1-\theta}} \right) \ge r_0^{\frac{1-2\theta}{1-\theta}} + \frac{2\theta - 1}{(1-\theta)\tau} k.$$

Consequently, $r_k \le ck^{\frac{1-\theta}{1-2\theta}}$, for some $c > 0$.
- If $\theta \in]0, 1/2]$, then $r_k \le \tau(r_{k-1} - r_k)$, $k = 1, 2, \ldots$ Therefore $r_k \le \frac{\tau}{\tau+1} r_{k-1}$, which implies right away $r_k \le cq^k$, where $q = \tau/(\tau+1)$ and $c = r_0$.
- If $\theta = 0$, then $\|x^k - x^{k+1}\| \ge 1/L$ for $x^k \ne x^{k+1}$ with k sufficiently large. The inequality

$$f(x^k) - f(x^{k+1}) \ge \frac{\rho}{2} \|x^k - x^{k+1}\|^2 \ge \frac{\rho}{2L^2}$$

for $x^k \neq x^{k+1}$ shows that DCA terminates after a finite number of iterations.

4 Applications to Trust-Region Subproblems and Quadratic Programs

Theorem 3.5 provides the convergence rate of DCA depending on the Łojasiewicz exponent at critical points of the objective function. Unfortunately, showing the existence of the Łojasiewicz exponent and computing it are, in general, two challenging problems. In this final section, we consider the two classes of usefully important DC programs: trust-region problems and quadratic programs with polyhedral convex constraints. The numerical computations for these two special problems were reported in the papers [4–6,20,23–26], which showed the high efficiency of the DCAs. We will prove that the Łojasiewicz exponent is $1/2$ for both DC programs and, as a result, state the Root-linear convergence rate of the DCAs applied to them.

4.1 Trust-Region Subproblems

Trust-region subproblems constitute a nice and useful class of important nonconvex programs, more exactly DC programs

$$\min \left\{ \frac{1}{2} x^T Q x + \langle b, x \rangle : \ x \in X, \ \|x\| \leq r \right\}, \tag{20}$$

or equivalently,

$$\min \left\{ f(x) := \frac{1}{2} x^T Q x + \langle b, x \rangle + \chi_C(x) : \ x \in X \right\}, \tag{21}$$

where Q is an $n \times n$ real symmetric matrix, $b \in X$, r is a positive scalar, and $C := \{x \in X : \|x\| \leq r\}$.

They play a key role in the trust-region method [27], which consists in solving a sequence of trust-region subproblems [2–5,28–30]. It can be considered as an improved Newton type method and is recognized to be among the most robust, stable and efficient method in nonlinear programming (see [27] where a chapter is devoted to DC programming and DCA for solving (TRSP)). It enjoys the following properties:

 (i) TRSP is one of rare nonconvex programs which possess verifiable global optimality conditions (quite close to KKT conditions). TRSP has only one local-nonglobal solution.

(ii) The set of KKT points for TRSP is contained in at most $2m + 2$ disjunctive subsets, where the objective function has the same value and m is the number of the distinct negative eigenvalues of the symmetric matrix defining the quadratic function of TRSP. These properties should promote inexpensive local descent methods which can perform finitely many restartings to converge to global solutions of TRSP. In [4,23], we investigated the DCA to solve TRSP. Thanks to its particular structure, the DCA is very simple: it consists of computing, at each iteration, the projection

of a point on a Euclidean ball, which is explicit and requires only matrix-vector products. In practice, the DCA converges to the global solution of TRSP. The inexpensive implicitly restarted Lanczos method of Sorensen is used to check the optimality of solutions provided by the DCA. When a nonglobal solution is found, a simple numerical procedure is introduced both to find a feasible point having a smaller objective value and to restart the DCA at this point. It is shown that in the nonconvex case, the DCA converges to the global solution of the trust-region subproblem, using only matrix-vector products and requiring at most $2m + 2$ restarts. Numerical simulations proved the robustness and efficiency of the DCA compared to related standard methods, especially for large-scale problems. It is also worth noting that TRSP plays an important role in BB (Branch-and-Bound) algorithms, using ellipsoidal outer approximation techniques for lower bounding, and this method has been applied successfully in many global algorithms [5,6,24–26].

The next theorem provides, under one suitable assumption, a Łojasiewicz exponent of some critical points of the objective function f and, as a result, the convergence rate of the DCA applied to (21). First, we recall the following lemma from ([31], Lemma 4.1).

Lemma 4.1 *Let Q be an $n \times n$ real symmetric matrix. Then, there exists a constant $M > 0$ such that the following inequality holds*

$$|\langle Qx, x\rangle|^{1/2} \leq M\|Qx\| \text{ for all } x \in X.$$

Consequently, we have:

Lemma 4.2 *Let $g(x):=\frac{1}{2}\langle x, Qx\rangle + \langle b, x\rangle$ be a quadratic function defined on X. If $\bar{x}$ is a critical point of g, that is, $\nabla g(\bar{x}) = Q\bar{x} + b = 0$, then there is $M > 0$ such that*

$$|g(x) - g(\bar{x})| \leq M\|Qx + b\|^2 \text{ for all } x \in X.$$

Proof As $\nabla g(\bar{x}) = Q\bar{x} + b = 0$, one has

$$g(x) - g(\bar{x}) = \frac{1}{2}\langle Q(x - \bar{x}), x - \bar{x}\rangle \text{ for all } x \in X.$$

The conclusion follows directly from Lemma 4.1 by noting $Qx + b = Q(x - \bar{x})$. □

Theorem 4.1 *Suppose $\bar{x} \in C$ is a critical point of the problem (21). That is, there exists $\bar{\lambda} \geq 0$ such that*

$$Q\bar{x} + b + \bar{\lambda}\bar{x} = 0, \quad \bar{\lambda}(\|\bar{x}\| - r) = 0. \tag{22}$$

In addition, assume that either $\|\bar{x}\| < r$ or $\|\bar{x}\| = r$ and

$$\langle v, Qv + \bar{\lambda}v\rangle > 0 \text{ for all } v \in X \text{ with } \langle \bar{x}, v\rangle = 0, \|v\| = 1. \tag{23}$$

Then, the Łojasiewicz exponent of the function f at the critical point $\bar{x}$ is $1/2$.

Proof Obviously, one has

$$\partial^F f(x) = \begin{cases} \{Qx + b + \lambda x : \lambda \geq 0\}, & \text{if } \|x\| = r, \\ \{Qx + b\}, & \text{if } \|x\| < r. \end{cases} \quad (24)$$

Let $\bar{x} \in C$ be a critical point of the function f satisfying (22). If $\|\bar{x}\| < r$, then $Q\bar{x} + b = 0$. According to Lemma 4.1, there exists $M > 0$ such that for all $x \in C$, one has

$$M\|Qx + b\| = M\|Q(x - \bar{x})\| \geq \frac{1}{2}|(x - \bar{x})^T Q(x - \bar{x})|^{1/2} = |f(x) - f(\bar{x})|^{1/2}.$$

Suppose now $\|\bar{x}\| = r$ such that (23) is satisfied. Then, there exists $\bar{\lambda} \geq 0$ such that $Q\bar{x} + b + \bar{\lambda}\bar{x} = 0$. By using again Lemma 4.1, we can find $M > 0$ such that for all $x \in C$, there holds

$$M\|Qx + b + \bar{\lambda}x\| = M\|Q(x - \bar{x}) + \bar{\lambda}(x - \bar{x})\|$$
$$\geq |(x - \bar{x})^T Q(x - \bar{x}) + \bar{\lambda}\|x - \bar{x}\|^2|^{1/2}.$$

On the other hand, the Taylor expansion gives

$$f(x) - f(\bar{x}) = \frac{1}{2}(x - \bar{x})^T Q(x - \bar{x}) + \langle Q\bar{x} + b, x - \bar{x} \rangle$$
$$= \frac{1}{2}(x - \bar{x})^T Q(x - \bar{x}) - \bar{\lambda}\langle \bar{x}, x - \bar{x} \rangle.$$

The last two relations imply

$$M\|Qx + b + \bar{\lambda}x\| \geq |f(x) - f(\bar{x}) + \bar{\lambda}(\|x\|^2 - \|\bar{x}\|^2)|^{1/2}. \quad (25)$$

Let $x \in C$ with $\|x\| < r$. If $\bar{\lambda} = 0$, then by (25)

$$M\|Qx + b\| \geq |f(x) - f(\bar{x})|^{1/2},$$

otherwise, $\bar{\lambda} > 0$, then for x sufficiently near to $\bar{x}$, one also has

$$M\|Qx + b\| \geq M\bar{\lambda}\|x\| - M\|Qx + b + \bar{\lambda}x\| \geq |f(x) - f(\bar{x})|^{1/2}.$$

Let us show that there exist $m, \varepsilon > 0$ such that for all $\lambda \in]\bar{\lambda} - \varepsilon, \bar{\lambda} + \varepsilon[, x \in B(\bar{x}, \varepsilon)$ with $\|x\| = r$,

$$\|Qx + b + \lambda x\| \geq m\|Qx + b + \bar{\lambda}x\|. \quad (26)$$

Indeed, assuming the contrary implies the existence of two sequences $\{x^k\}$ and $\{\lambda^k\}$ such that $x^k \to \bar{x}$, $\lambda_k \to \bar{\lambda}$ with $\|x^k\| = r$, $\lambda_k \neq \bar{\lambda}$, and $Qx^k + b + \bar{\lambda}x^k \neq 0$ for all

$k \in \mathbb{N}$ such that

$$\lim_{k \to \infty} \frac{\|Qx^k + b + \lambda_k x^k\|}{\|Qx^k + b + \bar{\lambda} x^k\|} = 0.$$

By setting $u^k = x^k - \bar{x}$, assume that $\left\{ \dfrac{u^k}{\|u^k\|} \right\}$ converges to v, then $\|v\| = 1$; $\langle \bar{x}, v \rangle = 0$, and one obtains

$$\lim_{k \to \infty} \frac{\|Qx^k + b + \lambda_k x^k\|}{\|Qx^k + b + \bar{\lambda} x^k\|} = \lim_{k \to \infty} \frac{\|Qu^k + \lambda_k u^k + (\lambda_k - \bar{\lambda})\bar{x}\|}{\|Qu^k + \bar{\lambda} u^k\|} = 0.$$

Hence, there exists $t \in \mathbb{R}$ such that $Qv + \bar{\lambda} v = t\bar{x}$. Therefore, $\langle v, Qv + \bar{\lambda} v \rangle = 0$, which contradicts (23).

By virtue of (25) and (26), one obtains, for some $L, \varepsilon > 0$,

$$L\|Qx + b + \lambda x\| \geq |f(x) - f(\bar{x})|^{1/2}, \ \forall \lambda \in (\bar{\lambda} - \varepsilon, \bar{\lambda} + \varepsilon);$$
$$\forall x \in B(\bar{x}, \varepsilon), \ \|x\| = r. \tag{27}$$

Finally, for any $\lambda \in \mathbb{R}$ with $|\lambda - \bar{\lambda}| \geq \varepsilon$ and for all x near to $\bar{x}$, one has

$$M\|Qx + b + \lambda x\| \geq M|\lambda - \bar{\lambda}|\|x\| - M\|Qx + b + \bar{\lambda} x\| \geq |f(x) - f(\bar{x})|^{1/2}.$$

The proof is completed. □

Remark 4.1 If the critical point $\bar{x}$ satisfies (23), then it is either a local-nonglobal minimizer or a global minimizer of the problem (21) (see, e.g., [4]).

4.2 Nonconvex Quadratic Programs with Polyhedral Convex Constraints

Finally, we consider nonconvex quadratic programs of the form

$$\min \left\{ \frac{1}{2} x^T Q x + \langle b, x \rangle : \ x \in K \right\}, \tag{28}$$

where K is a polyhedral convex set in X, defined by

$$K = \{ x \in X : \ \langle a_i, x \rangle \leq r_i, \ i = 1, \ldots, p \},$$

with $a_i \in X$, $r_i \in \mathbb{R}$, $i = 1, \ldots, p$.

This subclass of DC programs is well known in nonconvex programming and global optimization from both theoretical and algorithmic points of view [5]. It is known to be NP-hard except for some special cases [32].

The DCA was intensively investigated in our various works for (28) in the general case [5,6,20,21,24]. A quadratic function is an attractive example of a DC-function having various DC decompositions which enjoy great effects on resulting DCA schemes.

We developed in [20] suitable DC decompositions and corresponding DCAs for the general case of (28). The DCA then consists in the solution of a sequence of convex quadratic programs.

For solving large-scale problems, we have recourse to (polynomial-time) interior point methods for solving convex quadratic programs. In [5,6,21], we provided a new regularization technique based on DC programming and DCA to handle indefinite Hessians in a primal-dual interior point framework for nonconvex quadratic programming problems. The new regularization leads automatically to a strongly factorizable quasidefinite matrix in the Newton system. Numerical results for large-scale problems (up to 400,000 dimensions) showed the robustness and the efficiency of our approach compared with the powerful software LOQO [33].

As with trust-region subproblems, we will establish below, for the first time in the literature, the calculation of Łojasiewicz exponent and, consequently, the convergence rate of the DCA applied to (28).

According to (2), the problem (28) is rewritten equivalently,

$$\min\{f(x) = \frac{1}{2}x^T Q x + \langle b, x \rangle + \chi_K(x) : x \in X\}. \tag{29}$$

This problem can be formulated as a DC program by some DC decompositions. Various DCA-based algorithms have been proposed in [5,6,20,21,24,25].

Let $\bar{x} \in K$ be a critical point of the problem (29). That is,

$$0 \in \partial^F f(\bar{x}) = Q\bar{x} + b + N(K, \bar{x}), \tag{30}$$

where the normal cone $N(K, \bar{x})$ to K at $\bar{x}$ is simply calculated by

$$N(K, \bar{x}) = \left\{ \sum_{i=1}^{p} \lambda_i a_i : \ \lambda_i \geq 0, \ \lambda_i(\langle a_i, \bar{x} \rangle - r_i) = 0, \ i = 1, \ldots, p \right\}. \tag{31}$$

Theorem 4.2 *Let $\bar{x} \in X$ be a critical point of the problem (29). Then, there are $\tau > 0$ and $\delta > 0$ such that*

$$|f(x) - f(\bar{x})|^{1/2} \leq \tau \|x^*\| \quad \forall x \in B(\bar{x}, \delta), \ x^* \in \partial^F f(x).$$

Proof For $x \in K$, denote by

$$I(x) = \{i \in \{1, \ldots, p\} : \ \langle a_i, x \rangle - r_i = 0\}.$$

Let $\bar{\delta} > 0$ be such that

$$I(x) \subset I(\bar{x}) \quad \text{for all} \ x \in B(\bar{x}, \bar{\delta}). \tag{32}$$

For each $I \subset I(\bar{x})$, define the quadratic function $\Phi_I : X \times \mathbb{R}^{|I|} \to \mathbb{R}$ by

$$\Phi_I(x, \Lambda_I) = \frac{1}{2}\langle x, Qx \rangle + \langle b_i, x \rangle + \sum_{i \in I} \lambda_i(\langle a_i, x \rangle - r_i),$$

$$(x, \Lambda_I) \in X \times \mathbb{R}^{|I|}, \Lambda_I := (\lambda_i)_{i \in I}.$$

Then,

$$\nabla\Phi_I(x, \Lambda_I) = (Qx + b + \sum_{i \in I} \lambda_i a_i, (\langle a_i, x \rangle - r_i)_{i \in I}), (x, \Lambda_I) \in X \times \mathbb{R}^{|I|}.$$

For a given $I \subset I(\bar{x})$, denote by C_I the convex cone generated by $\{a_i : i \in I\}$:

$$C_I = \left\{ \sum_{i \in I} \lambda_i a_i : \lambda_i \geq 0, \ i = 1, \ldots, p \right\},$$

and consider the following two cases:

Case 1 $-(Q\bar{x} + b) \in C_I$. That is, there exists $\bar{\Lambda}_I = (\bar{\lambda}_i)_{i \in I} \in \mathbb{R}^{|I|}_+$ such that

$$Q\bar{x} + b + \sum_{i \in I} \bar{\lambda}_i a_i = 0.$$

It follows that $(\bar{x}, \bar{\Lambda}_I)$ is a critical point of $\Phi_I : \nabla\Phi_I(\bar{x}, \bar{\Lambda}_I) = 0$. According to Lemma 4.2, we find $M_I > 0$ such that

$$|\Phi(x, \Lambda_I) - \Phi(\bar{x}, \bar{\Lambda}_i)|^{1/2} \leq M_I \|\nabla\Phi(x, \Lambda_I)\| \ \forall (x, \Lambda_I) \in X \times \mathbb{R}^{|I|}. \quad (33)$$

Case 2 $-(Q\bar{x} + b) \notin C_I$. Since C_I is closed, one has

$$\inf_{\Lambda_I \in \mathbb{R}^{|I|}_+} \left\| Q\bar{x} + b + \sum_{i \in I} \lambda_i a_i \right\| = \varepsilon_I > 0.$$

Given $\delta_I > 0$ such that

$$\left| \frac{1}{2}(\langle x, Qx \rangle - \langle \bar{x}, Q\bar{x} \rangle) + \langle b, x - \bar{x} \rangle \right| < \varepsilon_I^2/4, \ \|Q(x - \bar{x})\| < \varepsilon_I/2 \ \forall x \in B(\bar{x}, \delta_I).$$

Then, for all $x \in K \cap B(\bar{x}, \delta_I) \ \Lambda_I = (\lambda_i)_{i \in I} \in \mathbb{R}^{|I|}_+$ with

$$\langle a_i, x \rangle - r_i = 0, \ \text{for } i \in I,$$

one has

$$|f(x) - f(\bar{x})|^{1/2} = \left| \frac{1}{2}(\langle x, Qx \rangle - \langle \bar{x}, Q\bar{x} \rangle) + \langle b, x - \bar{x} \rangle \right|^{1/2} \atop \leq \left\| Q\bar{x} + b + \sum_{i \in I} \lambda_i a_i \right\|. \quad (34)$$

Set

$$\tau = \max\{1, M_I : I \subset I(\bar{x})\} \text{ and } \delta = \min\{\bar{\delta}, \delta_I : I \subset I(\bar{x})\}.$$

Now let $x \in K \cap B(\bar{x}, \delta)$ and $x^* \in \partial^F f(x)$ be given. Then, by (32), $I := I(x) \subset I(\bar{x})$, and there is $\Lambda_I = (\lambda_i)_{i \in I} \in \mathbb{R}^{|I|}$ such that

$$x^* = Qx + b + \sum_{i \in I} \lambda_i a_i.$$

If Case 1 holds, then from (33), one has

$$|f(x) - f(\bar{x})|^{1/2} = |\Phi(x, \Lambda_I) - \Phi(\bar{x}, \bar{\Lambda}_i)|^{1/2}$$
$$\leq \tau \|\nabla \Phi(x, \Lambda_I)\| = \tau \|x^*\|.$$

If Case 2 holds, then by (34),

$$|f(x) - f(\bar{x})|^{1/2} \leq \left\| Q\bar{x} + b + \sum_{i \in I} \lambda_i a_i \right\| = \|x^*\|.$$

The proof is completed.   □

5 Conclusions

We have proved, in this paper, the convergence of the whole DCA sequence when the objective function is subanalytic and, at every critical point, at least one of the two DC components of the considered DC decomposition has locally Lipschitz derivative. We also established the convergence rate which depended on the Łojasiewicz exponent at a limit point. In particular, if this Łojasiewicz exponent is 1/2, then the DCA is Root-linearly convergent. In addition, we showed that, for trust-region problems, the Łojasiewicz exponent is effectively 1/2 at critical points satisfying the sufficient second order optimality condition, and for quadratic programs this exponent 1/2 holds at every critical point of the tackled DC program. Needless to say, the convergence rate of the DCA, related to both main classes of trust-region subproblems and quadratic programs, will allow to enhance its efficiency, by adapting DC decompositions to the specific structures of the studied DC programs with subanalytic data. Further developments will be devoted to computing the Łojasiewicz exponents for other classes of real-world DC programs.

Compliance with ethical standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Le Thi, H.A., Pham Dinh, T.: Large scale global molecular optimization from distance matrices by a DC optimization approach. SIAM J. Optim. **14**(1), 77–116 (2003)
2. Le Thi, H.A., Pham Dinh, T.: The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems. Ann. Oper. Res. **133**, 23–48 (2005)
3. Pham Dinh, T., Le Thi, H.A.: Convex analysis approach to DC programming: theory, algorithms and applications. Acta Math. Vietnam. **22**, 289–355 (1997)
4. Pham Dinh, T., Le Thi, H.A.: A DC optimization algorithm for solving the trust region subproblem. SIAM J. Optim. **8**(2), 476–505 (1998)
5. Le Thi, H.A.: DC programming and DCA. http://www.lita.univ-lorraine.fr/~lethi/index.php/dca.html (homepage) (2005)
6. Le Thi, H.A., Pham Dinh, T.: DC programming and DCA: theory, algorithms and applications. Special Issue of Math. Program. Series B, Dedicated to Thirty Years of Developments, 169(1) (2018)
7. Bolte, J., Daniilidis, A., Lewis, A.: Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamic systems. SIAM J. Optim. **17**(4), 1205–1223 (2007)
8. Łojasiewicz, S.: Sur le problème de la division. Studia Math. **18**, 87–136 (1959)
9. Łojasiewicz, S.: Une propriété topologique des sous-ensembles analytiques. In: Les Equations aux Dérivées Partielles, Editions du Centre National de la Recherche Scientifique, Paris, pp. 87–89 (1963)
10. Łojasiewicz, S.: Sur la géométrie semi- et sous-analytique. Ann. Inst. Fourier (Grenoble) **43**(5), 1575–1595 (1993)
11. Bierstone, E., Milman, P.: Semianalytic and subanalytic sets. IHES Publ. Math. **67**, 5–42 (1988)
12. Shiota, M.: Geometry of Subanalytic and Semialgebraic Sets. Progress in Mathematics, vol. 150. Birkhauser Boston, Inc., Boston, MA (1997)
13. Absil, P.-A., Mahony, R., Andrews, B.: Convergence of the iterates of descent methods for analytic cost functions. SIAM J. Optim. **16**, 531–547 (2005)
14. Attouch, H., Bolte, J.: The convergence of the proximal algorithm for nonsmooth functions involving analytic features. Math. Program. **116**, 5–16 (2009)
15. Attouch, H., Bolte, J., Svaiter, B.F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward backward splitting, and regularized Gauss Seidel methods. Math. Program. **137**, 91–129 (2013)
16. Mahey, P., Oualibouch, S., Pham Dinh, T.: Proximal decomposition on the graph of monotone operator. SIAM J. Optim. **5**(2), 454–466 (1995)
17. Mordukhovich, B.S.: Variational analysis and generalized differentiation. I. Basic theory. Grundlehren Math. Wiss, vol. 330. Springer, Berlin (2006)
18. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
19. Rockafellar, R.T., Wets, J.-B.: Variational Analysis. Springer, New York (1998)
20. Le Thi, H.A., Pham Dinh, T.: Solving a class of linearly constrained indefinite quadratic problems by DC algorithms. J. Global Optim. **11**, 253–285 (1997)
21. Pham Dinh, T., Le Thi, H.A., Akoa, F.: Combining DCA and interior point techniques for large-scale nonconvex quadratic programming. Optim. Method Softw. **23**(4), 609–629 (2008)
22. Hiriart-Urruty, J.-B., Lemarchal, C.: Convex Analysis and Minimization Algorithms II., Advanced Theory and Bundle Methods, 2nd edn. Springer, Berlin (1996)
23. Pham Dinh, T., Le Thi, H.A.: Difference of convex function optimization algorithms (DCA) for globally minimizing nonconvex quadratic forms on Euclidean balls and spheres. Oper. Res. Lett. **19**(5), 207–216 (1996)
24. Le Thi, H.A., Pham Dinh, T., Muu, L.D.: A combined D.C. optimization-ellipsoidal branch-and-bound algorithm for solving nonconvex quadratic programming problems. J. Comb. Optim. **2**(1), 9–28 (1998)
25. Le Thi, H.A., Pham Dinh, T.: A branch-and-bound method via D.C. optimization algorithm and ellipsoidal technique for box constrained nonconvex quadratic programming problems. J. Global Optim. **13**(2), 171–206 (1998)
26. Le Thi, H.A.: An efficient algorithm for globally minimizing a quadratic function under convex quadratic constraints. Math. Program. **87**(3), 401–426 (2000)
27. Conn, A.R., Gould, N.I.M., Toint, PhL: Trust-Region Methods. MPS/SIAM Ser. Optim. SIAM, Philedalphia (2000)
28. Tuan, H.N.: Convergence rate of the Pham Dinh-Le Thi algorithm for the trust-region subproblem. J. Optim. Theory Appl. **154**(3), 904–915 (2012)

29. Le Thi, H.A., Pham Dinh, T., Yen, N.D.: Behavior of DCA sequences for solving the trust-region subproblem. J. Global Optim. **53**, 317–329 (2012)
30. Tuan, H.N., Yen, N.D.: Convergence of Pham Dinh-Le Thi's algorithm for the trust-region subproblem. J. Global Optim. **55**(2), 337–347 (2013)
31. Huynh, V.N., Théra, M.: Error bounds for systems of lower semicontinuous functions in Asplund spaces. Math. Program. **116**, 397–427 (2009)
32. Vavasis, S.A.: Nonlinear Optimization: Complexity Issues. Oxford University Press, Oxford (1991)
33. Vanderbei, R.J.: LOQO: an interior point code for quadratic programming. Optim. Method Softw. **11**(1–4), 451–484 (1999)

Affiliations

Hoai An Le Thi[1] (ID) · Van Ngai Huynh[2] · Tao Pham Dinh[3]

Van Ngai Huynh
ngaivn@yahoo.com

Tao Pham Dinh
pham@insa-rouen.fr

[1] Laboratory of Theoretical and Applied Computer Science, University of Lorraine, Site de Metz, Ile du Saulcy, 57045 Metz, France

[2] Department of Mathematics, University of Quynhon, 170 An Duong Vuong, Qui Nhon, Vietnam

[3] Laboratory of Mathematics, INSA-Rouen, University of Normandie, Avenue de l'Université, 76800 Saint-Étienne-du-Rouvray Cedex, France